EXPLORING USERS’ VISUAL IMPRESSION OF A JAPANESE STREETSCAPE BY CORRELATING ATTENTION WITH SPEECH

Utilizing eye-tracking technology for computer-aided architectural planning

KHANG NGUYEN-TRAN1, YUVAL KAHLON2, TAKUYA OKI3, NAYA MARSATYASTI4, RYO MURATA5 and HARUYUKI FUJII6
1,2,3,4,5,6Tokyo Institute of Technology.
khang.ntr@gmail.com 0000-0002-0535-0179
kahlon.y.aa@m.titech.ac.jp 0000-0002-3914-7945
oki.t.ab@m.titech.ac.jp 0000-0002-4848-0707
naya.marsatyasti@gmail.com 0000-0001-9654-9222
murata.r.ac@m.titech.ac.jp 0000-0003-4010-2178
fujii.h.aa@m.titech.ac.jp 0000-0002-8984-8512

Abstract. Considering users’ subjective impressions is a challenging question in architectural design. Answering this question is a human-centred approach which is critical for designing spaces that afford true well-being for their users relating to Sustainable Development Goal 3, or SDG 3: “Good Health and Well-being”). While various methods for evaluating users’ subjective impressions exist, their focus is mainly on classifying spaces into pre-determined categories representing the users’ impressions. Such classification fails to capture the richness of users’ actual interpretation, as reflected in descriptions provided by users when they express freely their impression about architectural space. Aiming to capture this richness, we extend the current state-of-the-art methods and propose an integrated approach to extract and analyse the users’ interpretations while interviewing them observing a Tokyo streetscape. In addition, by comparing their comments and gaze, we expose correlations between abstract descriptions and the architectural space, which are difficult to obtain using existing methods. Our insights are a stepping-stone for enhancing computer-aided architecture by integrating visual impressions into the design processes.

Keywords. Visual Impression; Gaze Analysis; Japanese Architecture; Oku; Well-being; SDG 3.

1. Introduction

In the rapidly evolving research into visual impressions of architectural space (VIAS), an important question is how to consider people's subjective impressions when designing architectural space. Since people can effectively express their subjective interpretation in natural language, it is essential to understand how to relate linguistic
expressions with architectural spaces. Currently, the field is led by quantitative approaches that often use methods to match visual stimuli with linguistic descriptions, with or without numerical scales. While these approaches have shed light on how we experience architectural spaces, they are limited in their ability to extract and precisely describe users' VIAS. One reason for this is their reliance on pre-determined categories, which subjects are requested to choose from in reporting VIAS. Such pre-determined categorization is bound to miss some of the subtleties and richness which characterize VIAS. Recently, the research on VIAS has taken a new approach using state-of-the-art technologies for eye tracking (Oki and Kizawa, 2021). This approach yields valuable information regarding the subject's attention, which is necessary for developing a comprehensive understanding of individual reports of VIAS. Our work aims to further the efforts beyond the previous findings by integrating them with a qualitative analysis of verbal reports. The findings are expected to contribute to a new approach for studying VIAS and could be integrated into computer-aided architecture design. Furthermore, considering how complexly the visual representation of architectural space can influence human perception, the architectural design process must involve a hidden but crucial sense concerning people. This study's findings can offer a unique method to provide insights that decrypt the user's interpretation of architectural space and allow designers to configure spaces that ensure and promote human well-being in urban spaces and cities. This contribution to the well-being of inhabitants via a human-centred design approach accords with the Sustainable Development Goal 3 (SDG 3) to “Ensure healthy lives and promote well-being for all at all ages.”.

2. Background

2.1. EXTRACTING VISUAL IMPRESSION OF STREETSCAPES

Concerning current work on extracting VIAS, various methods have been developed. The Semantic Differential method is commonly used in research to correlate the subject's opinion with the stimuli in photographs or the actual space as in the studies of the impression of streetscapes based on indications such as the sky factor (Nishio and Ito, 2015), the green ratio (Nakamura et al., 2010), or the perceived attractiveness of urban spaces via the relation between the distance from skyscrapers and VIAS (Wada and Kishimoto, 2011). Moreover, impression evaluation using Affinity Diagrams also provided insights into differences in perception of different age groups, genders, and cultural backgrounds (Kacha et al., 2015).

In recent years, eye-tracking devices have become available at a lower price, which allows designers and researchers to use this technology more frequently in studies attempting to evaluate VIAS. For instance, a study analysed the relationship between landscape order, gaze tendency and impression factors (Ohno, 2018). Concerning urban streetscapes, a study analysed how urban street edges affected pedestrians' field of vision and how spatial factors impacted this relation (Simpson et al., 2019). Developing an objective way to evaluate VIAS, our previous experiment examined the relationship between the subjects' gazing tendency and their evaluation of streetscape attractiveness (Oki and Kizawa, 2021) by comparing the gaze tendencies and visual saliency estimated by a deep learning model. However, how to get insight from such data is still in its infancy, and the interpretability of eye measurement results also
remains an issue. Building upon these efforts, this study aims to obtain better comprehensive descriptions of VIAS through speech and attention and retrace them in the visual representation of the streetscape. Also, to reflect the subjects’ impression with a high degree of reliability, we combined the benefits offered by gaze analysis with the diligence given by the long-standing method of protocol analysis (Jiang and Yen, 2010) into a framework correlating subjective understanding and interpretation of VIAS.

2.2. THE CONCEPT OF OKU FOR JAPANESE STREETSCAPE

The Japanese concept of ‘Oku’, literally translated as 'innermost space' (Maki, 2018, p.156), often used to describe spatial formation in Japan. This concept signifies relative distance obtained from a visual permeability through a sequence of screens, planes, or filters toward the core of space. Although understood as 'spatial depth' (Monnai, 2009, p.183) or 'multilayers space' (Jonas, 2011, p.99), Oku is not limited in the narrow sense of physical distance but is often associated with a subjective sense of feeling ‘moving into unknown places’ (Kohte, 2017, p.25) or evoking a ‘sense of curiosity’ (Totten, 2016, p.258). This abstract concept cannot be understood using a single, all-encompassing definition but is also manifested in different architectural approaches seeking to express the sense of Oku via the manipulation of spatial boundaries and sequences. For example, in his book Exterior Design in Architecture, Ashihara mentioned it in the discussion about the hidden order of how to break the spatial monotony and create a sense of profundity by the 'ingenious use of the inside and outside corners' (Ashihara, 1970, p. 94-95). Also, Ando incorporates this concept in his masterpiece The Church of the Light to an articulation of 'sentiment-fundamental space' (Ando, 1995, p.74) that incites the visitor to seek symbolism in horizontal depth. While this notion of Oku is an essential characteristic of dwellings or spiritual spaces somehow related to the idea of 'hidden', 'inner', or 'secret' (Maki, 2018, p.157), when applied to streetscapes, it represents a particular character of Japanese cities. Discussions regarding the overlapping external layers of building facades, the gap in topography, and the volume of objects (Maki, 2018, p.135-148) are essential to understand this spatial formation. Outside of Japan, similar concepts existed in western literature using definitions such as 'here and there' (Cullen, 1961, p.182), 'phenomenal transparency' (Rowe, 1963, p.46), 'soft edges' (Gehl, 2011, p.183-184), etc. It is clear that the richness of Oku presents a potentially wide range of interpretations from the designer and users and is, therefore, a suitable topic for this study.

3. Method

3.1. DATA COLLECTION

The case studies were collected from the Daikanyama streetscape with Hillside Terrace, a masterpiece of Maki and an exceptional representation of the concept Oku. This is where buildings, streets, public spaces are composed of elements that allow multiple visual penetrations. To understand the visual impression mechanism, we introduced a combined method of data collection using photo segmentation, eye tracking, and interviews in three steps according to the framework in Figure 1.
First, in the process, 13 photos of the streetscape in the Daikanyama area were collected using a Fujifilm XT-10 digital camera with a setting closest to a human viewpoint. Following the classification of elements based on their 'environmental meaning' (Rapoport, 1982, p.87), photos were segmented into five types of elements such as vertical (building parts), horizontal (pavement), artificial shape (street furniture), natural shape (vegetation), and others (people, sky, background scenery, etc.). Second, we ran an experiment in a laboratory environment by projecting the original photos on a screen while interviewing nine subjects. The subjects from various countries are trained in the architectural design field and represent a variety of nationalities. We observed them as they shared their impressions about Oku from the photos and collected data regarding their comments and gazes. During the experiment, a screen-based eye tracker (Tobii Pro nano) is used to collect data on the subject's gazes. The data collection settings are described in Figure 1.

Figure 1. Framework and process of study

<table>
<thead>
<tr>
<th>Data collection</th>
<th>Target: Oku that can be seen or felt from photos of Tokyo streetscape</th>
</tr>
</thead>
<tbody>
<tr>
<td>PHOTOS</td>
<td>GAZE</td>
</tr>
<tr>
<td>-----</td>
<td>----</td>
</tr>
<tr>
<td>Gaze distribution</td>
<td>Time density</td>
</tr>
<tr>
<td>Visual Attention</td>
<td>Visual Composition</td>
</tr>
</tbody>
</table>

Step 1: Take photos  
Step 2: Experiment on subject's attention & speech  
Step 3: Correlate results and findings

Figure 2. Data collection
3.2. DATA PROCESSING AND ANALYSIS

We extracted subjects' interpretations from the protocols, as well as their gaze data at the corresponding times in the experiment, then integrated them into diagrams. A deep examination by comparing verbal descriptions with the subjects' attention enables us to identify the connections and the deviations between the different types of data, thus yielding important insights about VIAS. It should be noted that during the data collection and processing, invalid data were removed due to calibration issues.

3.2.1. Protocol Analysis

Transcriptions were used to identify all moments when the concept of Oku is discussed by each subject, which we defined as 'an event'. In each event, we first marked the time when the subject began speaking about Oku, then extracted the subject’s interpretation of the concept of Oku, which is expressed in their utterances regarding VIAS. Finally, if applicable, the degree of Oku was included with the interpretation (e.g., 'strong feeling' of Oku). These interpretations were then classified into those in which the subject associated Oku with a physical architectural element, such as a 'longwall', and others in which the concept was associated with abstract qualities of the space, such as 'mysterious atmosphere'. These classifications served as the basis for understanding how the subject related the concept of Oku with the actual or visual space in each photo.

3.2.2. Gaze Analysis

By using a small screen-based eye tracker, the subjects' gaze coordinates and pupil diameter were recorded with a 1/60-second cycle. Although saccades, in which the gaze moves quickly, are included in the gaze behaviour data, in this study, we analysed gazing behaviour in which the subjects were looking at objects. Specifically, we defined attention 'as a state in which the eye movement speed is less than 10 degrees/second for more than 100 milliseconds' (Fukuda et al., 1996, p. 203) based on the distribution of the measured eyes movement speed to extract the attention points. Then, we created a kernel density distribution of attention points for each conversation by subject and photo. From this distribution, we visualized the subjects' attention as heat maps reflecting their visual behaviour during the event. Besides, attention heat maps were generated to capture the gaze behaviour in the minute before and after an event. This step resulted in an image set showing their visual focus, which accompanied their speech during the event. Finally, by extracting the visual attention in the heat maps during the moment that Oku was mentioned or discussed, we listed the physical parts of the streetscape, which were focused on in each event.

4. Results and Discussions

4.1. GAZE TENDENCIES DURING INTERPRETATION OF OKU

We have extracted six photos that were indicated among all subjects as possessing the quality of Oku. For each of these, we display a heat map reflecting the subject's gaze during discussion of Oku, thus forming the 5x6 matrix in Figure 3. This matrix can be used to search for gaze tendencies among the subjects. For instance, by examining
photo number 9 in terms of similarities, five out of six subjects were attracted to the vanishing point. This observation seems reasonable, as this point represents the deepest visible area. Notwithstanding, S6 gazed at an elongated element on the ground belonging to the foreground. This observation poses a challenge for the previous explanation of Oku relating to a relative distance concerning depth. Considering that the element points to the vanishing point, it seems reasonable to assume that Oku is experienced since the vanishing point is implied. In other words, the directionality of an element can lead our eyes into the depths of the space and result in an experience of Oku. Contrary to the above, the subject's explanation revealed another aspect of interpretation. He mentioned that the element frames a 'small space' for sitting, which he associated with 'the Japanese tea house', thus rendering an impression of Oku. This episode demonstrates the value in the analysis of VIAS by integrating gaze data with qualitative analysis of speech, which can reveal deep insights about one's views.

<table>
<thead>
<tr>
<th></th>
<th>Gazing time density during conversation about Oku</th>
<th>Low</th>
<th>Medium</th>
<th>High</th>
</tr>
</thead>
<tbody>
<tr>
<td>Subject</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td></td>
<td>4</td>
<td>5</td>
<td>9</td>
</tr>
<tr>
<td>5</td>
<td></td>
<td>11</td>
<td>13</td>
<td></td>
</tr>
<tr>
<td>9</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>11</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>13</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Figure 3. Gazing tendency in subject's response during conversations about Oku

4.2. ESTIMATING FREQUENCY AND VARIETY OF INTERPRETATION VIA SPEECH AND ATTENTION

By collecting the words subjects used to talk about Oku and presenting them along with gaze time density in heat maps, we can see the variety of VIAS across several subjects, as shown in Figure 4. These contain various interpretations of Oku, not only as an abstract concept but also as related to physical elements and their perception (for example, for S1, Oku is described by 'building's part' linked to 'barrier', while S9 linked it to a sense or a feeling that can or cannot be seen).
EXPLORING USER'S VISUAL IMPRESSION OF A JAPANESE STREETSCAPE BY CORRELATING ATTENTION WITH SPEECH

Furthermore, by superimposing the probability distribution obtained from kernel density estimation and segmentation images as shown in Figure 5, we can estimate which elements were being gazed at. This estimation enables us to raise hypotheses regarding the variety of elements in the formation of VIAS, both individually and collectively. For instance, vegetation showing in green colour seems to catch a very small portion of S4's and S7's attention. Meanwhile, the prevalence of tall blue columns indicates a collective tendency to gaze at building façades more than other elements.

4.3. TRACING SOURCES OF VIAS

After extracting the heat maps and utterances reflecting subjects' attention and thoughts when discussing the concept of Oku, we combined them into concise yet informative episodes in Figure 6. For each subject, three attention heat maps are presented: before discussing Oku, during the initial report on their impression, and immediately after that moment. The initial instance refers to the moment in time which follows the interviewer's question regarding Oku, which naturally draws their attention to this concept (note that, in some cases, a subject began speaking of Oku without being prompted to discuss it). Relevant time durations appear at the bottom of each heat map. Each heat map is accompanied by speech data, where important phrases are marked in a bold font. The interviewer is represented by the letter 'I' and subjects are represented by the letter 'S'. Finally, the subject's VIAS regarding Oku and its potential source are summarized at the bottom left of each episode. For example, consider S1, who reported that the space has a ‘…type of Oku’. The subject tied this to the existence of a staircase, which then creates a 'barrier'. Thus, Oku is interpreted here as related to the existence of barriers, which finds its source in the architectural elements of the staircase.
Figure 6. Interpretation of Oku in speech with the gaze of each subject
Besides, by setting heat maps side-by-side with the relevant explanations provided by the subjects, we can have a better understanding of subjective interpretation. For example, how S1 above has traced the impression of Oku to the existence of a 'barrier', and that element is caused by the fact of the 'staircase in the centre'. A question that could be raised is how does the staircase create a sense of 'barrier'? To answer it, we examine what S1 was referring to when talking about a 'staircase'. Consulting the accompanying heat maps reveals that the subject was focused on a small area at the end of the central staircase while ignoring most of the staircase's mass. Also, the focused area coincides with the staircase end, which is the visual boundary limit between it and what lies beyond it. Furthermore, notice that before the question regarding Oku, S1 was focusing on the child playing at the centre of the staircase. This observation strengthens the assumption that his focus on the edge was prompted by the question and thus is tied with his subjective interpretation of Oku. Other similar in-depth examinations are possible by checking the other episodes. For example, in the case of S7, we can relate the subject's interpretation of Oku as the ability to see through elements. In the attention heat map, the focus is on a specific area of the pergola (i.e., the wooden planks hung on the outdoor terrace) and do not obstruct the view.

5. Conclusion and Future Work

This work explored VIAS on the topic of Oku. Based on aspects related to the streetscape, we have qualitatively assessed users' impressions containing speech and attention. The findings provided insights into an approach that decrypts the complexity of the user's interpretation. First, it allows us to access the tendency and variety of interpretations of tangible aspects in spatial design via the combination of gaze and photo segmentation. Second, it enhances the ability to decrypt the subjective understanding of spatial formation via the proposed method of tracing the source of VIAS. Applying this method has proved useful in offering insights into subjective impressions observed when discussing Oku based on the tendency of attention. Consequently, the existing theories of VIAS can be enriched via this human-centred approach. For example, Okabe's related concept of 'subjective depth' (Okabe, 2017, p.158) can now be related with empirical findings concerning VIAS and thus be enriched or even reassessed. From a broader perspective, similar experiments can be reconstructed, moving beyond streetscapes and the concept of Oku in studying VIAS.

As for the scope and limitations of this study, the number of samples is limited. While the proposed method showed a certain degree of reliability when correlating the interpretation with the spatial design abstract concept, further investigation is necessary to understand how subjects construct impressions based on interpretation. Also, using more samples could further provide enough data to build a deep learning model describing the people's gaze movement or attention. Such a model will enable us to evaluate the attractiveness of places as a computer-aided design simulation tool that integrates into the design process. Finally, the insights collected from this research represent a unique direction for exploring the hidden mechanism of the user's perception. The findings will help designers enhance the design processes for urban assessment that involve users' evaluations. Access to this intangible dimension of architectural space is essential for urban dwellers' well-being, which corresponds to the aim of indication SDG 3 and is a step forward to designing sustainable cities.
References